

MOTE-NAS: Multi-Objective Training-based Estimate for Efficient Neural Architecture Search

Yu-Ming Zhang¹ Jun-Wei Hsieh² Xin Li³ Ming-Ching Chang³
Chun-Chieh Lee¹ Kuo-Chin Fan¹
¹National Central University ³University at Albany
²National Yang Ming Chiao Tung University

Reporter: Yu-Ming Zhang

Two Types of NAS Methods

- ***Macro Perspective***

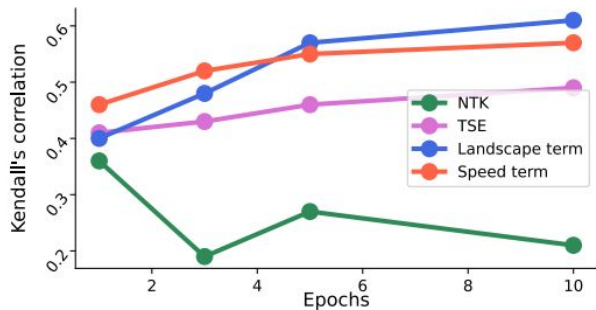
- Capturing non-convex nature by **mathematical view**
- **NTK-based** estimates (*K-NAS, TE-NAS, Eigen-NAS, ...*)
- ...

- ***Micro Perspective***

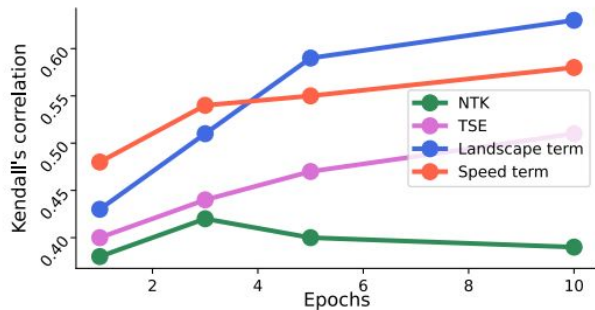
- Capturing non-convex nature by **gradient changes**
- **Gradient-based** estimates (*Snip, SynFlow, TSE, ...*)
- ...

The Problem of Neural Tangent Kernel (NTK)

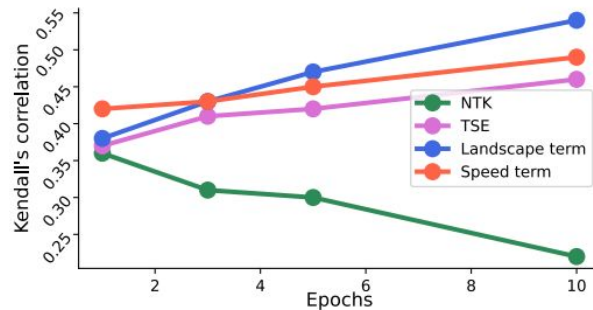
- NTK is not consistent after training, violates its basic assumptions
- TSE performs even better than NTK



(a) CIFAR-10



(b) CIFAR-100

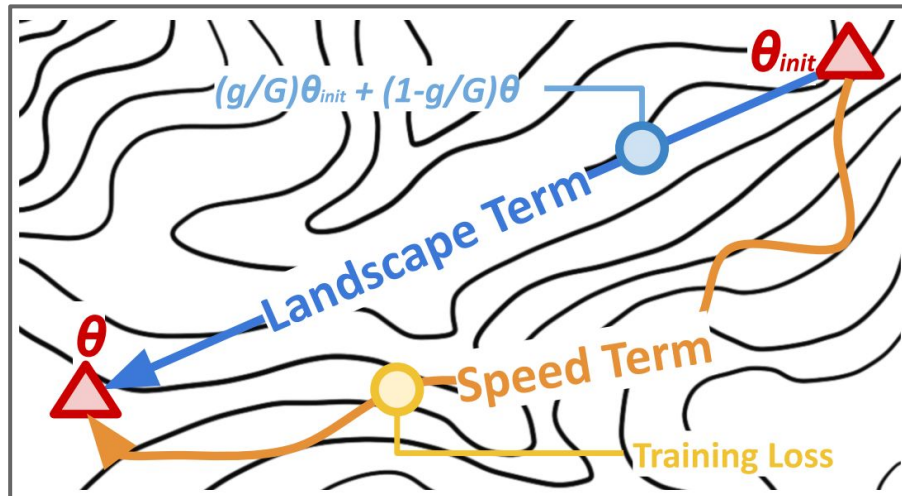


(c) ImageNet-16

Two Terms of MOTE

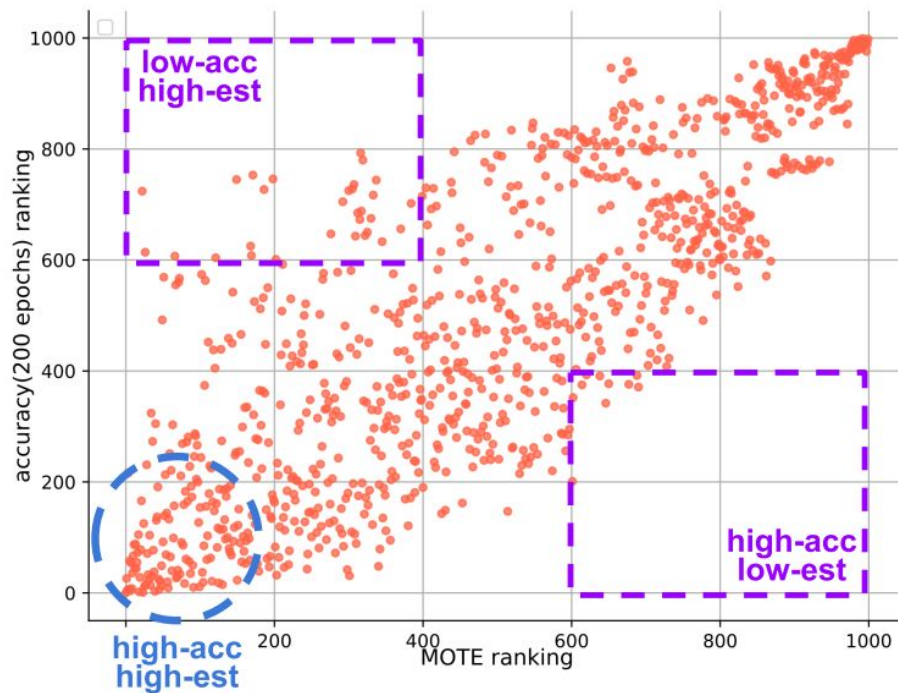
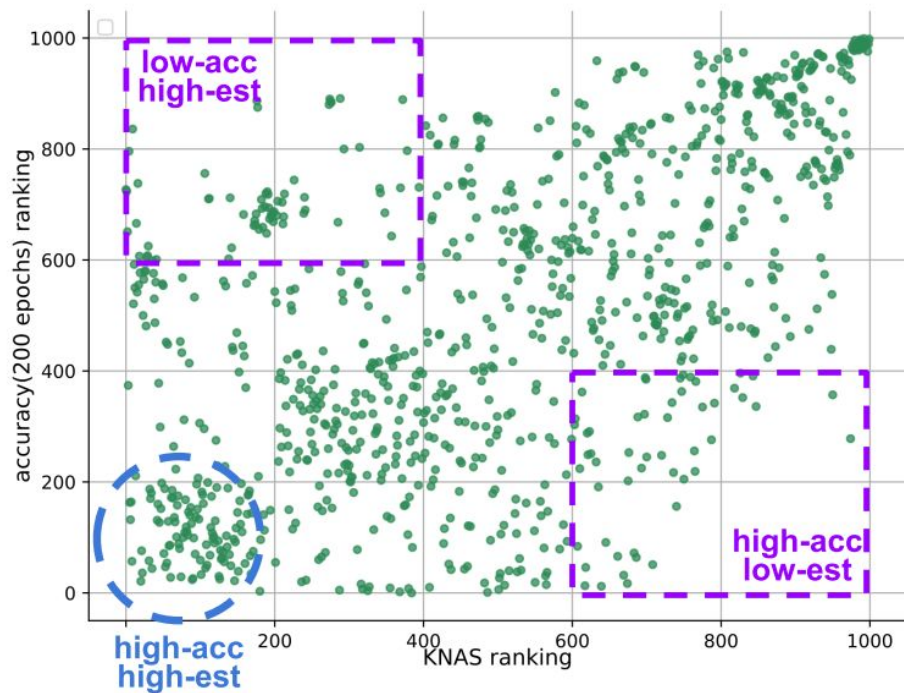
- Landscape term is used to capture the non-convex at a macro view
- Speed term is used to capture the non-convex at a micro view

$$MOTE = \underbrace{f\left(\sum_{g=0}^G \mathcal{J}_{\theta(g)}\right)}_{\text{landscape term}} + \underbrace{f\left(\sum_{e=1}^E \frac{l_e}{t_e}\right)}_{\text{speed term}}$$

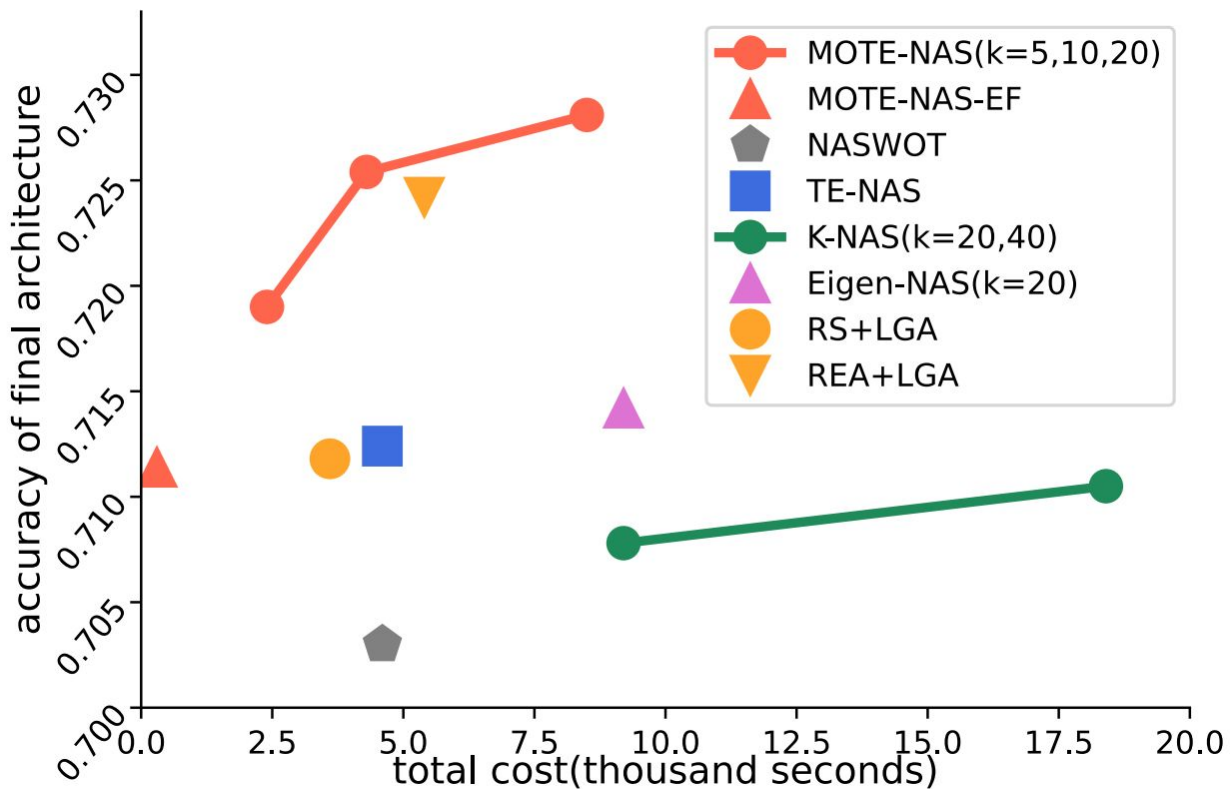


sketch map of loss landscape

MOTE-NAS Compared to K-NAS(*NTK-based*)



MOTE-NAS Compared to Others



Thank You for Listening